REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
22-06-2001

2. REPORT DATE type
final

3. DATES COVERED *(From - To)*
01NOV97 - 31OCT00

4. TITLE AND SUBTITLE
Planning-Based Information Agents

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N000-98-1-0147

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Daniel S. Weld, Professor
Department of Computer Science & Engineering
University of Washington

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Washington
3935 University Way NE, Box 355754
Seattle, WA 98195

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dept. of the Navy, Office of Naval Research
Seattle Regional Office
1107 NE 45th Street, #350
Seattle, WA 98105-4631

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release, distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

20010705 054

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Carol Zuiches
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPONE NUMBER** *(Include area code)*
U	UU	U			206-543-4043

Final Report for "Planning-Based Information Agents"

Grant number N00014-98-1-0147
November 1, 1997 - October 31, 2000.

Daniel S. Weld
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195
weld@cs.washington.edu

SUMMARY

Networked information systems are making so much data available that people can't find it themselves. Software agent technology promises to amplify human decision making capabilities by gathering information from disparate sources in parallel and integrating it in real time. However, in order to make today's prototype systems realize their potential, several bottlenecks must be overcome. First, information-gathering agents need robust and efficient execution so they can process large data sets, cope with network failure and site outage. Secondly, in order to scale to the level of thousands of information sources, agents need algorithms for locating information sources; automatically creating wrappers for those sources, processing XML based representations of those sources, and routing queries to the appropriate sources.

PROGRESS

We've formulated the problem of wrapper induction, proved some theoretical PAC bounds on the performance of such systems, devised a number of learning algorithms that solve the problems for different classes of sources, implemented the algorithms, and performed empirical tests on the implementations. Many others have extended our seminal results.

We've built a prototype system that automatically identifies, classifies, wraps, and query routes to over ten thousand specialized information sources. Key ideas include two novel methods for query routing: intelligent probing of CGI scripts to determine their expertise and using the Yahoo categorization of specialized information sources as a kind of semantic networks.

We've built the MULDER system which takes natural language questions, parses them, composes a set of Internet search engine queries of differing specificity using novel paraphrasing technology, sends the queries to an engine such as Google, downloads likely pages returned by Google, parses regions of the resulting pages, extracts candidate answers to the original questions, and votes to determine which are the most likely correct answer(s). MULDER outperforms commercial systems such as Google and AskJeeves. Ablations studies show the benefit derived by each of our techniques.

We've developed extensions to the proposed W3C standard XML query language allowing for updates to XML documents. We have implemented a dozen different update methods on a variety of relational encodings of XML data, and performed experiments to determine which methods work best. We've implemented a highly optimized execution system for data integration. The resulting system, Tukwila, can handle four orders of magnitude more data than its predecessor system Razor. Key ideas include adaptivity at all levels of the architecture, interleaved planning and execution, and a novel double-pipelined join algorithm which greatly reduces latency when combining data from sources connected via low or medium-speed networks.

We've extended Tukwila to natively handle semi structured, XML information. Our algorithms leverage existing database technology yet incorporates novel query processing operators (such as XScan). Detailed empirical experiments show that our methods vastly outperform previous methods.

We've extended planning technology to handle interleaved query planning and execution as well as traditional AI planning in the context of uncertainty.

We've built two new planning systems. TGP is a temporal planner that uses Graphplan-like mutual exclusion reasoning to achieve impressive performance. LPSAT compiles resource planning problems into a combined linear-programming/propositional satisfiability representation, which is then solved using a novel combination of incremental simplex and Davis-Putnam systematic SAT algorithms.

Finally, we've implemented the Tiramisu web site management system. Tiramisu separates the design of a web site from its implementation, allowing the use of multiple implementation tools while supporting a high-level declarative model of the site.

ACCOMPLISHMENTS

Design, implementation and test of next-generation, scalable, fully autonomous wrapper creation system.

Implementation of prototype web resource detector and query routing system.

Implementation and testing of Tukwila adaptive execution system for information integration.

Design, implementation and experimentation on MULDER, the first fully automated question-answering system for the WWW.

Design of XML update language, comparative implementation of update methods and experimental evaluation.

Experiments showing utility of double pipelined join, interleaved planning and execution, and other Tukwila features.

Design, implementation and experimentation on conformant graphplan.

Design, implementation and experimentation on contingent graphplan.

Design, implementation and experimentation on factored expansion graphplan.

Design, implementation and experimentation on TGP temporal planner.

Design, implementation and experimentation on LPSAT resource planner.

Design and implementation of Tiramisu web-site management system.

TRANSITIONS

My primary collaborators are Professor Oren Etzioni and Professor Alon Halevy, both at the University of Washington, and Dr. David Smith at NASA Ames Research Center. Our work on wrapper induction has been adopted and extended by Professor Nick Kushmerick (former student) now at Dublin City University, Ireland, by Dr. Steve Minton and Dr. Craig Knoblock at ISI, and by the group of Professor Tom Mitchell at CMU. Nimble technology (a startup company which I co-founded with Professor Halevy) has licensed the Tukwila data integration system. NASA is interested in fielding our planning work. The W3C standards body is considering incorporating our XML update methods in the next standard.

AWARDS

I was made AAAI Fellow for my "significant contribution to the development of qualitative reasoning methods, software agent technology, and plan synthesis algorithms."

I was presented with the WRF / TJ Cable Endowed Professorship.

PUBLICATIONS

Ives, Z., Levy, A., Tatarinov, I., and Weld, D. "Updating XML," *2001 ACM Conference on Management of Data* (SIGMOD-01), Santa Barbara, CA, May 2001.

Kwok, C., Etzioni, O., and Weld D. "Scaling Question Answering to the Web," *Tenth International World Wide Web Conference* (WWW10), Hong Kong, May 2001.

Levy, A. and Weld, D., "Intelligent Internet Systems," *Artificial Intelligence*, **118** (1--2): 1--14, April 2000.

Wolfman, S. and Weld, D. "Combining Linear Programming and Satisfiability Solving for Resource Planning," *Knowledge Engineering Review* 15:1, 2000.

Wolfman, S. and Weld, D. "The LPSAT Engine & its Application to Resource Planning," Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99), Stockholm, Sweden, August 1999.

Smith, D. and Weld, D. "Temporal Planning with Mutual Exclusion Reasoning," Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99), Stockholm, Sweden, August 1999.

Ives, Z. and Florescu, D. and Friedman, M. and Levy, A. and Weld, D. "An Adaptive Query Execution System for Data Integration," 1999 ACM Conference on Management of Data (SIGMOD-99), Philadelphia, PA, June 1999.

Lau, T., Etzioni, O., and Weld, D. "Privacy Interfaces for Information Management," Communications of the ACM, 1999.

Anderson, C. A. and Levy, A. Y. and Weld, D. S. "Declarative web-site management with Tiramisu," Proceedings of the Second International Workshop on The Web and Databases (WebDB '99), 1999.

Lau, T., Etzioni, O., and Weld, D. "Privacy Interfaces for Information Management," *Communications of the ACM*, October 1999.

Lau, T. and Weld, D., "Programming by Demonstration: An inductive learning formulation," 1999 ACM International Conference on Intelligent User Interfaces (IUI-99), Orlando, FL, January 1999.

Weld, D. and Anderson, C. and Smith, D., "Extending Graphplan to Handle Uncertainty & Sensing Actions," *Fifteenth National Conference on Artificial Intelligence* (AAAI-98), 9 pages, Madison, WI, July 1998.

Smith, D. and Anderson, C. and Weld, D., "Contingent Graphplan," *Fifteenth National Conference on Artificial Intelligence* (AAAI-98), 9 pages, Madison, WI, July 1998.

Anderson, C. and Smith, D. and Weld, D., "Conditional Effects in Graphplan," *Fourth International Conference on Artificial Intelligence Planning Systems* (AIPS-98), 9 pages, Pittsburgh, PA, June 1998.

Etzioni, O., Golden, K. and Weld, D., "Sound and Efficient Closed-World Reasoning for Planning," *Artificial Intelligence*, **89**:113--148, 1997.

Weld, D. and Anderson, C. and Smith, D., "Extending Graphplan to Handle Uncertainty & Sensing Actions," *Fifteenth National Conference on Artificial Intelligence* (AAAI-98), 9 pages, Madison, WI, July 1998.

Smith, D. and Anderson, C. and Weld, D., "Contingent Graphplan," *Fifteenth National Conference on Artificial Intelligence* (AAAI-98), 9 pages, Madison, WI, July 1998.

Anderson, C. and Smith, D. and Weld, D., “Conditional Effects in Graphplan,” *Fourth International Conference on Artificial Intelligence Planning Systems* (AIPS-98), 9 pages, Pittsburgh, PA, June 1998.

Kushmerick, N. and Doorenbos, R. and Weld, D., “Wrapper Induction for Information Extraction,” *Fifteenth International Joint Conference on Artificial Intelligence* (IJCAI-97), 7 pages, Nagoya Japan, August 1997.

Friedman, M. and Weld, D., “Efficient Execution of Information Gathering Plans,” *Fifteenth International Joint Conference on Artificial Intelligence* (IJCAI-97), 7 pages, Nagoya Japan, August 1997.

Ernst, M. and Millstein, T. and Weld, D., “Automatic SAT-Compilation of Planning Problems,” *Fifteenth International Joint Conference on Artificial Intelligence* (IJCAI-97), 8 pages, Nagoya Japan, August 1997.

REPORT OF INVENTIONS AND SUBCONTRACTS

(Pursuant to "Patent Rights" Contract Clause) (See Instructions on back)

Form Approved
OMB No. 9000-0095
Expires Aug 31, 2001

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (9000-0095), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR COMPLETED FORM TO THIS ADDRESS. RETURN COMPLETED FORM TO THE CONTRACTING OFFICER.

1.a. NAME OF CONTRACTOR/SUBCONTRACTOR University of Washington	c. CONTRACT NUMBER N00014-98-1-0147	2.a. NAME OF GOVERNMENTPRIME CONTRACTOR	c. CONTRACT NUMBER	3. TYPE OF REPORT *(X one)* a. INTERIM ☒ b. FINAL
b. ADDRESS *(Include ZIP Code)* 3935 University Way NE, Box 355754 Seattle, WA 98195-	d. AWARD DATE *(YYYYMMDD)* 19971101	b. ADDRESS *(Include ZIP Code)*	d. AWARD DATE *(YYYYMMDD)*	4. REPORTING PERIOD *(YYYYMMDD)* a. FROM 19971101 b. TO 20001031

SECTION I - SUBJECT INVENTIONS

5. "SUBJECTINVENTIONS"REQUIREDTO BE REPORTEDBY CONTRACTOR/SUBCONTRACTOR *(If "None," so state)*

NAME(S) OF INVENTOR(S) *(Last, First, Middle Initial)* a.	TITLE OF INVENTION(S) b.	DISCLOSURE NUMBER, PATENT APPLICATION SERIAL NUMBER OR PATENT NUMBER c.	ELECTION TO FILE PATENT APPLICATIONS *(X)* d. (1) UNITED STATES (a) YES	(1) UNITED STATES (b) NO	(2) FOREIGN (a) YES	(2) FOREIGN (b) NO	CONFIRMATORY INSTRUMENT OR ASSIGNMENT FORWARDED TO CONTRACTING OFFICER *(X)* e. (a) YES	(b) NO
	None							

f. EMPLOYER OF INVENTOR(S) NOT EMPLOYED BY CONTRACTOR/SUBCONTRACTOR

(1) (a) NAME OF INVENTOR *(Last, First, Middle Initial)*

(b) NAME OF EMPLOYER

(c) ADDRESS OF EMPLOYER *(Include ZIP Code)*

(2) (a) NAME OF INVENTOR *(Last, First, Middle Initial)*

(b) NAME OF EMPLOYER

(c) ADDRESS OF EMPLOYER *(Include ZIP Code)*

g. ELECTED FOREIGN COUNTRIES IN WHICH A PATENT APPLICATION WILL BE FILED

(1) TITLE OF INVENTION

(2) FOREIGN COUNTRIES OF PATENT APPLICATION

SECTION II - SUBCONTRACTS *(Containinga "Patent Rights" clause)*

6. SUBCONTRACTSAWARDEDBY CONTRACTOR/SUBCONTRACTOR *(If "None," so state)*

NAME OF SUBCONTRACTOR(S) a.	ADDRESS *(Include ZIP Code)* b.	SUBCONTRACT NUMBER(S) c.	FAR "PATENT RIGHTS" d. (1) CLAUSE NUMBER	(2) DATE *(YYYYMM)*	DESCRIPTION OF WORK TO BE PERFORMED UNDER SUBCONTRACT(S) e.	SUBCONTRACT DATES *(YYYYMMDD)* f. (1) AWARD	(2) ESTIMATED COMPLETION
None							

SECTION III - CERTIFICATION

7. CERTIFICATIONOF REPORTBY CONTRACTOR/SUBCONTRACTOR *(Not required if: (X as appropriate))* □ SMALL BUSINESSor □ NONPROFITORGANIZATION

I certifythat the reportingparty has proceduresfor promptidentification and timelydisclosureof "Subject Inventions," that such procedureshave beenfollowed and that all "Subject Inventions" have been reported.

a. NAME OF AUTHORIZED CONTRACTOR/SUBCONTRACTOR OFFICIAL *(Last, First, Middle Initial)* Zuiches, Carol	b. TITLE Director, Grant & Cont. Svcs.	c. SIGNATURE SINH SIMMONS GRANT & CONTRACT MANAGER ACTING FOR CAROL ZUICHES	d. DATE SIGNED 6/26/01

PREVIOUS EDITION MAY BE USED.

Date: Tue, 26 Jun 2001 16:46:51 -0700 (PDT)
From: Alicen L. Smith <asmith@cs.washington.edu>
To: collins5@u.washington.edu
Subject: Re: onr final report (fwd)

---------- Forwarded message ----------
Date: Fri, 22 Jun 2001 13:28:41 -0400
From: Daniel Weld <weld@dakobed.com>
To: "'asmith@cs.washington.edu'" <asmith@cs.washington.edu>
Subject: Re: onr final report

No patents
Thanks

Dan

-----Original Message-----
From: Alicen L. Smith <asmith@cs.washington.edu>
To: Dan Weld <weld@cs.washington.edu>
Sent: Fri Jun 22 09:47:49 2001
Subject: Re: onr final report

There may be a few more questions, but so far all I need to know is:

> Did you acquire any patents as a result of this research?

If so, please list:

> Inventors' names
> title of invention(s)
> patent (application) number

Thanks,
Alicen